

# Dense motion estimation using regularization constraints on local parametric models

Ioannis Patras, *Member, IEEE*, and Marcel Worring, *Member, IEEE*, and Rein van den Boomgaard

**Abstract**—This paper presents a method for dense optical flow estimation in which the motion field within patches that result from an initial intensity segmentation is parametrized with models of different order. We propose a novel formulation which introduces regularization constraints between the model parameters of neighboring patches. In this way, we provide the additional constraints for very small patches and for patches whose intensity variation cannot sufficiently constrain the estimation of their motion parameters. In order to preserve motion discontinuities we use robust functions as a regularization mean. We adopt a three frame approach and control the balance between the backward and forward constraints by a real-valued direction field on which regularization constraints are applied. An iterative deterministic relaxation method is employed in order to solve the corresponding optimization problem. Experimental results show that the proposed method deals successfully with motions large in magnitude, motion discontinuities and produces accurate piecewise smooth motion fields.

**Index Terms**—Motion Estimation, Regularization, Intensity Segmentation, Robust Regression

## I. INTRODUCTION

Estimating dense optical flow fields in unknown scenes has always been problematic due to the fact that the motion estimation problem is ill-posed [26]. Over the years a number of researchers attempted to overcome the ill-posedness by imposing a variety of constraints on the spatial or temporal coherency of the motion field [19]. Block-based motion estimators assume that the motion within rectangular blocks follows a simple, most often translational, parametric model. Regularization techniques assume a globally [15] or piecewise [6] smooth motion field. Segmentation-driven methods [17] assume that the scene can be decomposed into a relatively small number of regions such that the motion of each region can be described by a simple parametric model.

The block-based and the global smoothness based approaches obviously make unrealistic assumptions about the structure of the motion field. On the other hand regularization and segmentation based approaches are faced with the non trivial problem of determining automatically the region of support on which the coherency constraints should be imposed. The realization that by relying on motion information alone it is very difficult to obtain good localization of the region of support has steered a number of researchers to hybrid intensity/motion-based approaches. In Markov Random Field formulations this is achieved by adapting the clique potential to the presence of an intensity edge [13]. Similarly, [20] and [1] adjust the smoothness constraint depending on the magnitude and the direction of the image gradient. However, the smoothness constraints they impose are weak in comparison to

parametric constraints. Other approaches [11] utilize an initial intensity segmentation in order to apply smoothness [29] or parametric constraints within each intensity segment. Such approaches have given promising results for motion-based segmentation but do not address inter-segment constraints. This imposes an unnatural limitation to the extent of the coherency region; usually regions with coherent motion extend beyond the borders of a single intensity segment. Furthermore, the estimation of the model parameters becomes difficult if the intensity segment is small, especially in the presence of motion with large magnitude. This poses an initialization problem to methods that iteratively a) merge intensity segments (e.g. [9]) based on their motion parameters and b) reestimate the motion parameters at the union of the merged regions. In order to overcome this problem Black and Jepson [5] utilize a dense optical-flow field to obtain an initialization of the motion parameters of the initial segments. However, such an initialization depends on the quality of the initial motion field which could be low at motion discontinuities or at areas with low intensity variation.

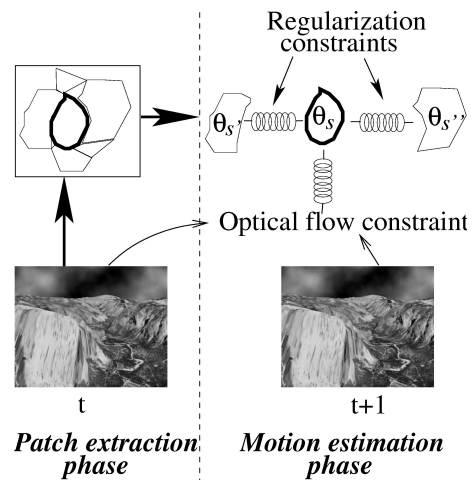


Fig. 1. Outline of the proposed method

In [23] we presented a method that addresses inter-segment constraints in the context of motion-based *segmentation*. In this paper, an early version of which appears at [24], in the context of motion *estimation*, we propose a framework that exploits the benefits of both the pixel-based robust regularization methods and the region-based motion estimation methods. In a first phase (Fig. 1) we apply an intensity-based segmentation to decompose the current frame into a number of intensity segments (hereafter called patches). In the second phase (Fig. 1), we

treat each patch as a site in an iterative relaxation scheme that estimates simultaneously the parametric models that describe the motion of the patches. In contrast to the methods in the literature, we address inter-segment constraints by applying robust regularization in the space of the motion parameters. This can sufficiently constrain the estimation of the parameters even for very small patches, provide coherent parameters for neighboring patches and at the same time preserve the motion discontinuities.

The remainder of the paper is structured as follows. In Section II we formulate the motion estimation as an optimization problem and in Section III describe the optimization procedure. In Section IV we present experimental results and finally in Section V conclusions are drawn.

## II. PROBLEM FORMULATION

### A. Patch Models

The essence of our method is that the problem of estimating a dense motion field is formulated as the problem of estimating a number of local parametric models, each of which describes the motion field at a local image patch. These patches are extracted at the first phase of our method (Fig. 1) by performing an initial intensity-based segmentation. The current frame, is first simplified with morphological operators (opening and closing by reconstruction) and, subsequently, pixels whose intensity differs less than a threshold  $T$  are grouped together. This segmentation method, which can be replaced by any other similar method, attempts to decompose the current frame into patches whose edges do not violate motion discontinuities and, at the same time, group in the same patch pixels that by themselves yield unreliable motion constraints due to low intensity variation around them. While a single frame intensity-based segmentation cannot, in general, respect motion discontinuities it has long been argued [5] [13] [20] [1] [11] [29] [9] that it can be very useful since a) in many scenes, especially in man-made environments, motion discontinuities usually do coincide with intensity edges and b) intensity edges can be much easier and better localized than motion discontinuities. In practice, the goals of a) not violating motion discontinuities and b) deriving patches with sufficient intensity variation are conflicting and the balance between them is controlled by the size of the morphological operators and the threshold  $T$ . In our scheme, which degrades gracefully to pixel-based methods when the patches contain a single pixel, we use a very conservative initial segmentation ( $1 \leq T \leq 3$ ) typical results of which are depicted in Fig. 4 and Fig. 5. Let us also note that our method does not consider the break up of patches at a later stage. Therefore, if the initial intensity contains patches which strive over motion discontinuities or, in general, patches whose motion cannot be described by the parametric models that we use, the motion field within the patches in question will be also partially erroneous.

Let us denote with  $S$  the set of patches  $s$  that are extracted by the initial intensity-based segmentation, with  $\Omega(s)$  the set of pixels  $x$  in patch  $s$  and with  $N_s$  the set of neighbors of the patch  $s$ . Finally, let us denote with  $\Omega(s, s')$  the set of pixels

along the border between neighboring  $s$  and  $s'$  and define  $b_{ss'} \equiv |\Omega(s, s')|$  as the corresponding common border length.

We introduce strong constraints for the pixels that belong to the same patch by assuming that the motion field  $u_x$  (where  $x$  is a pixel) within each patch  $s$  can be described by an (unknown) parametric model  $\theta_s$ . That is

$$u_x(\theta_s) = m_s^T(x)\theta_s, \quad x \in \Omega(s) \quad (1)$$

where  $m_s(x)$  is the motion model matrix that relates (linearly) the motion parameters  $\theta_s$  of the patch  $s$  to the motion vector  $u_x$  at the pixel  $x$ . The motion model matrix for the affine model is, for example:

$$m_s^T(x) = \begin{bmatrix} 1 & i(x) & j(x) & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & i(x) & j(x) \end{bmatrix} \quad (2)$$

where  $i(x)$  and  $j(x)$  are the coordinates of pixel  $x$ .

We handle patches of different size and shape by allowing models of different order (depending on the size and on the shape of the patches). Given the usual maximum patch size that our intensity segmentation yields, the highest order of the model that we use is the affine model. Lower order models are assigned to certain patches  $s$  by restricting certain parameters of  $\theta_s$  to be zero. These constraints are applied in the estimation phase as described in Section III-C.

Finally, we address occlusions by considering correspondences for each pixel  $x$  in both temporal directions (backward and forward) and to a different degree in each direction. The degree to which correspondences are sought in each direction is encoded in an unknown field  $O = \{o_x \in [0 \dots 1] : x \in G\}$ , where with  $G$  we denote the set of pixels in the image grid. Therefore,  $o_x$  varies between 0 and 1 to the degree that data constraints are derived from the previous and the next frame respectively. The direction field  $O$  is introduced in order to derive valid data (i.e. photometric) constraints in the areas where these are mostly needed; near motion discontinuities. Indeed, motion discontinuities create always occlusions, let these be in the next frame, in the previous frame or, in the worst case, at both. Classical motion estimators keep the direction fixed (e.g. backward) and address the occlusions implicitly or explicitly [25] as data outliers. In this way at occlusions a) the data constraints (i.e. constraints derived from establishing correspondences between the current and the previous frame) are not valid and b) it is unlikely that the regularization constraints (i.e. constraints derived from the motion field estimated in the neighborhood) give correct cues since it is unknown in advance where the discontinuity lies. As a result, in the estimated motion field, the discontinuity is placed within the occluded area. In contrast, the introduction of the direction field allows for valid data terms as long as valid correspondences exist at least in one of the neighboring frames. Therefore, it provides an elegant solution for occlusions either in the previous or in the next frame which are, by far, the ones most often encountered. Another three frame approach, which chooses the direction (backward or forward) with the larger cross-correlation match, is adopted by [10]. In contrast to our approach it makes binary decisions and is incorporated in a tensor-based framework. Finally, let us here

denote with  $O_s$  the direction field at all pixels that belong to segment  $s$ , that is  $O_s = \{o_x : x \in \Omega(s)\}$ .

### B. Motion Estimation as an Optimization Problem

We seek the minimization of a cost function with respect to the unknowns  $\Theta = \{\theta_s : s \in S\}$  and  $O = \{o_x : x \in G\}$ . We incorporate our work in the robust regularization framework and let the cost function consist of a data term that expresses the dependencies between the data and the unknowns and two regularization terms that express the interdependencies between the unknowns. More specifically the cost function is defined as follows:

$$C(\Theta, O) = \lambda_d \sum_{s \in S} f_d(\theta_s, O_s, s) + \lambda_c \sum_{s \in S} \sum_{s' \in N_s} f_c(\theta_s, \theta_{s'}, s, s') + \lambda_o \sum_{x \in G} \sum_{x' \in N_x} f_o(o_x, o_{x'}) \quad (3)$$

where  $\lambda_d$ ,  $\lambda_c$  and  $\lambda_o$  are constants whose ratio controls the relative importance of the data term, the motion regularization term and the direction regularization term. In what follows, we explain these terms.

1) *Data term*: The data term (first term of eq. (3)) expresses how consistent the motion parameters  $\Theta$  and the direction field  $O$  are with respect to the observed image intensities. It is defined as the summation of local data terms,  $f_d(\theta_s, O_s, s)$ , each one of which is defined on the basis of a patch  $s$ . Each local data term expresses how well the patch in question can be reconstructed, given the image intensities in the previous and in the next frame and an estimation of the unknowns that are related to patch  $s$  (i.e.  $\theta_s$  and  $O_s$ ). Thus, in the local data term,  $f_d(\theta_s, O_s, s)$ , we encode the evidence the image data provide us about the motion of patch  $s$ . More specifically,

$$f_d(\theta_s, O_s, s) = \sum_{x \in \Omega(s)} \rho(r_d(\theta_s, o_x, x)) \quad (4)$$

where  $\rho(r)$  is a robust error function [16] which in our experiments was taken to be the Lorentzian (i.e.  $\rho(r) = \log\left(1 + \frac{1}{2}\left(\frac{r}{\sigma}\right)^2\right)$ ) and

$$r_d(\theta_s, o_x, x) = o_x r_d^f(\theta_s, x) + (1 - o_x) r_d^b(\theta_s, x) \quad (5)$$

is the data residual at point  $x$ . It is defined as a linear combination of the forward ( $r_d^f(\theta_s, x)$ ) and backward ( $r_d^b(\theta_s, x)$ ) motion-compensated intensity differences at pixel  $x$ . The forward motion-compensated intensity differences is  $r_d^f(\theta_s, x) = I(x + u(\theta_s, x), t + 1) - I(x, t)$  (and similarly  $r_d^b(\theta_s, x) = I(x, t) - I(x - u(\theta_s, x), t - 1)$ ). Note that  $o_x$  controls the balance between the backward/forward data constraints at pixel  $x$ . The hope is that  $o_x$  can be estimated simultaneously with the motion parameters so that its value reflects the degree to which pixel  $x$  is visible at the previous and at the next frame.

2) *Motion regularization term*: An optimization with respect to the data term alone results in a region-based motion estimation scheme similar to, for example, the work of Odobez and Bouthemy [21]. However, the implicit assumption that the data term can provide sufficient and correct constraints cannot be guaranteed due to classical problems in motion estimation, such as the aperture problem and occlusions. Such problems

can deteriorate the motion estimation even of large patches but are particularly acute when the size of the patch is small.

The motion regularization term (second term of eq. (3)) provides the additional constraints that are essential in the absence of sufficient or reliable data constraints. It introduces interdependencies between the motion parameters  $\theta_s$  and  $\theta_{s'}$  of neighboring patches  $s$  and  $s'$  by penalizing motion parameters that are dissimilar. We do so by defining the motion regularization residual  $r_c(\theta_s, \theta_{s'}, \Omega(s, s'))$  as the discrepancy between the motion fields generated by  $\theta_s$  and  $\theta_{s'}$  in an area  $\Omega(s, s')$ . Similar measures have been used in the context of motion-based region merging by, for example, Gelgon and Bouthemy [11]. Clearly, small values of the regularization residual  $r_c(\cdot)$  indicate a smooth motion field, while large values of  $r_c(\cdot)$  indicate motion discontinuities. The latter are tolerated by the use of a robust function  $\rho(\cdot)$ . More specifically, the local costs  $f_c(\theta_s, \theta_{s'}, s, s')$ 's that comprise the regularization term are defined as follows

$$f_c(\theta_s, \theta_{s'}, s, s') = b_{ss'} \rho(r_c(\theta_s, \theta_{s'}, \Omega(s, s'))) \quad (6)$$

where the use of factor  $b_{ss'}$ , which is the length of the common border between patches  $s$  and  $s'$ , implies that in our formulation the larger the common border between neighboring patches the larger the penalty we introduce. Finally,  $\rho(\cdot)$  is a robust function and  $r_c(\theta_s, \theta_{s'}, \Omega(s, s'))$  the regularization residual. The latter, is defined as follows

$$r_c(\theta_s, \theta_{s'}, \Omega(s, s')) = \sqrt{\frac{1}{|\Omega(s, s')|} \sum_{x \in \Omega(s, s')} \|u_x(\theta_s) - u_x(\theta_{s'})\|^2} \quad (7)$$

where the role of  $|\Omega(s, s')|$  is to scale the motion residual  $r_c$  so that the same robust function  $\rho(\cdot)$  can be used at eq. (6) for all patches  $s$ . Using some algebra and eq. (1) we can express the motion residual as

$$r_c(\theta_s, \theta_{s'}, \Omega(s, s')) = \sqrt{(\theta_s - \theta_{s'})^T M_{ss'} (\theta_s - \theta_{s'})} \quad (8)$$

where

$$M_{ss'} = \frac{1}{|\Omega(s, s')|} \sum_{x \in \Omega(s, s')} m_{ss'}(x) m_{ss'}^T(x) \quad (9)$$

Eq. (8) is a compact form of eq. (7) which expresses the square of the motion residual as a quadratic function of the motion parameters. This allows easy differentiation with respect to the motion parameters and will be used in all subsequent derivations.

3) *Direction field regularization term*: The third term in eq. (3) imposes regularization constraints on the direction field  $O$  itself by introducing interdependencies between the values of the direction field at neighboring pixels. This aims at obtaining a piecewise smooth direction field which is in accordance to where the correspondences lie in image sequences depicting moving opaque objects. Regularization constraints are particularly important for the iterative optimization scheme that we introduce, particularly at the first iterations when the motion parameters are not near their true values. Ignoring the regularization constraints at this point might lead to a direction field that points at arbitrary directions, thus, deprive us from

valid data constraints. The direction field regularization term is a classical pixel-based regularization term in which the local costs  $f_o(o_x, o_{x'})$  are defined as a function of the residual between the values of the field in two neighboring pixels  $x$  and  $x'$ . More specifically,

$$f_o(o_x, o_{x'}) = \rho(o_x - o_{x'}) \quad (10)$$

An important issue in the patch-based regularization scheme that we propose is the balance between the data and the regularization constraints which is controlled by  $\lambda_d$ ,  $\lambda_c$  and  $\lambda_o$ . More specifically, the ratio  $\frac{\lambda_c}{\lambda_d}$  controls the smoothness of the motion field while the ratio  $\frac{\lambda_o}{\lambda_d}$  the smoothness of the direction field.

Looking at the local costs to which each set of motion parameters  $\theta_s$  participates, it becomes apparent that the data term is of the order of the size of the patch  $s$  (i.e.  $\sum_{x \in \Omega(s)} \rho()$ ), while the motion regularization terms are of the order of the perimeter of the patch ( $\sum_{s' \in N_s} b_{ss'} \rho()$ ). Their ratio depends naturally on the patches shape but, in general, the larger the patch the larger the ratio between the data and the motion regularization term. Thus, the larger the ratio  $\frac{\lambda_c}{\lambda_d}$  the smaller the patches for which a normalization between the data and the motion regularization terms is achieved. Roughly, the choice of the ratio  $\frac{\lambda_c}{\lambda_d}$  should be in the order of magnitude of the square root of the size of the segments for which such a normalization is desired.

Looking at the local costs to which each  $o_x$  participates it becomes apparent that the ratio between the local data and direction regularization terms are of constant order, that is of  $\frac{1}{\|N_x\|}$ . This is a classical regularization term and in all of our experiments we have chosen for values of large values of  $\lambda_o$  in order to obtain a rather smooth direction field.

Finally, let us note that our formulation is a generalization of pixel-based regularization methods and region-based motion estimation methods. Indeed, both can be derived from eq. (3) with the appropriate choices for the parameters  $\lambda_c$ , for the motion models  $m_s$  and for the initial patches. With  $\lambda_c = 0$  and fixed direction field  $O$  we derive a classical region-based motion estimation scheme such as the one of Odobez and Bouthemy [21]. With each patch containing a single pixel, a translational motion model  $m_s$  for each patch/pixel and a fixed direction field  $O$  we derive a classical pixel-based regularization scheme similar to the one that was introduced by Black and Anandan [6]. Furthermore, let us note that our formulation bears similarities with the method of Memin and Perez [18] in which local parametric models have been also proposed. However, their work relies on motion information itself for detecting motion discontinuities and does not consider backward/forward correspondences.

### III. OPTIMIZATION

The minimization of eq. (3) is a multidimensional non-linear optimization problem of a function with interdependencies between the unknowns and local minima. Such problems are traditionally solved using iterative deterministic (e.g. [6]) or stochastic (e.g. [12]) methods. For computational efficiency we adopt a deterministic approach which iterates through three

main stages as these are outlined in Table I. The first stage is the optimization with respect to the direction field. At this stage we visit each pixel  $x$  and optimize the cost function with respect to  $o_x$  keeping all the other parameters frozen. The optimization with respect to  $o_x$  is achieved by solving an equivalent weighted Least Squares problem. The second stage consists of making a linear approximation of the data residual with respect to the motion parameters. Finally, the third stage is the optimization with respect to the motion parameters. At this stage we visit each patch  $s$  and optimize the cost function with respect to  $\theta_s$  keeping all the other parameters frozen. The optimization with respect to  $\theta_s$  is achieved by solving an equivalent weighted Least Squares problem. Our scheme iterates a fixed amount of times within the third stage.

In order to overcome local minima and to estimate motions large in magnitude we a) incorporate our method in the multiscale framework using a Gaussian image pyramid [2] and b) reduce gradually (at each iterations  $k$ ) the scale parameters  $\sigma_d$  and  $\sigma_c$  in the robust function  $\rho()$ . The latter is common practice in reconstruction problems that involve preservation of discontinuities and is in the same lines with the ‘‘Graduated Non Convexity’’ algorithm by Blake and Zisserman [8]. Let us note here, that the choice of the values of the scale parameters depends on the robust function  $\rho()$  and on our choice of the value  $\tau$  above which a residual should be consider as an outlier. A good discussion can be found in [6] which derives that for the Lorentzian the scale parameter  $\sigma$  should be chosen as  $\sigma = \tau/\sqrt{2}$ . In our experiments we adopted a linear reduction scheme [6].

For a fixed number of iterations
A: Visit each pixel $x$ Optimize cost w.r.t. $o_x$ by solving an equivalent weighted LS problem
B: Make a linear approximation of the data residuals with respect to the motion parameters
C: For a fixed number of iterations Visit each patch $s$ Optimize cost w.r.t. $\theta_s$ by solving an equivalent weighted LS problem Goto C
Reduce $\sigma_d$ and $\sigma_c$ Go to A

TABLE I  
OUTLINE OF THE OPTIMIZATION SCHEME

In what follows we will explain in more detail the different stages of our optimization scheme.

#### A. Optimization with respect to the direction field

At the minimization of the cost function with respect to the  $o_x$  at iteration  $k+1$  we keep all the other unknowns frozen as these were estimated in the previous iteration. Then each site  $x$  is visited and updated. Formally, we seek  $o_x$  that minimizes the local cost:

$$L_o(o_x) = \lambda_d \rho(r_d(\theta_s^k, o_x, x)) + \lambda_o \sum_{x' \in N_x} \rho(o_x - o_{x'}^k) \quad (11)$$

$$= \lambda_d \rho(o_x r_d^f(\theta_s^k, x)) + (1 - o_x) r_d^b(\theta_s^k, x) + \lambda_o \sum_{x' \in N_x} \rho(o_x - o_{x'}^k) \quad (12)$$

under the constraint  $0 \leq o_x \leq 1$ .

We make a substitution that comes from the field of robust statistics and turns the optimization of eq. (11) into a weighted Least Squares problem. In schemes like ours, the minimization of a cost function that is the summation of “robust” terms  $\rho(r)$  (w.r.t. the residuals  $r$ ) can be shown to be equivalent (for a detailed discussion see [7] [14]) to a scheme which at iteration  $k + 1$  minimizes the summation of weighted squares  $l^{k+1}r^2$  of the residuals  $r$ . The equivalence between the minimization schemes holds for weights  $l^{k+1}$  that are defined as  $l^{k+1} = \frac{\rho'(r^k)}{2r^k}$ . Since  $l^{k+1}$  is in fact a function of  $r^k$ , in the subsequent derivations we will use the notation  $l(r^k)$  when referring to it. Finally, let us note that the above derivations are valid only for robust functions  $\rho(\cdot)$  that satisfy certain conditions [7]. In our experiments  $\rho(r)$  was taken to be the Lorentzian function (i.e.  $\rho(r) = \log\left(1 + \frac{1}{2}\left(\frac{r}{\sigma}\right)^2\right)$ ).

Then, the equivalent weighted Least Squares problem is the minimization, with respect to  $o_x$ , of

$$\lambda_d l_d (r_d(\theta_s^k, o_x^k, x)) r_d(\theta_s^k, o_x^k, x)^2 + \lambda_o \sum_{x' \in N_x} l_o (o_x^k - o_{x'}^k) (o_x - o_{x'}^k)^2 \quad (13)$$

under the constraint

$$0 \leq o_x \leq 1. \quad (14)$$

Eq. (13) is quadratic with respect to  $o_x$  therefore of the form

$$\alpha o_x^2 + \beta o_x + \gamma, \quad (15)$$

where the coefficients  $\alpha$ ,  $\beta$  and  $\gamma$  can be obtained by substituting eq. (5) in eq. (13) and after some algebra. Omitting the derivations and the functional dependencies for notational simplicity, we give the coefficients as:

$$\alpha = \lambda_d l_d (r_d^f - r_d^b)^2 + \lambda_o \sum_{x' \in N_x} l_o \quad (16)$$

$$\beta = 2\lambda_d l_d r_d^b (r_d^f - r_d^b) - 2\lambda_o \sum_{x' \in N_x} l_o o_{x'}^k \quad (17)$$

$$\gamma = \lambda_d l_d (r_d^b)^2 + \lambda_o \sum_{x' \in N_x} l_o (o_{x'}^k)^2 \quad (18)$$

The quadratic eq. (15) has a global minimum at  $-\frac{\beta}{2\alpha}$ . It can be easily shown (we omit the proof) that under the constraint of eq. (14) the global minimum is at

$$o_x^{k+1} = \max \left\{ 0, \min \left\{ 1, -\frac{\beta}{2\alpha} \right\} \right\}. \quad (19)$$

### B. Linear approximation of the data residual

At the second stage of our method we make an approximation of the data residual that is linear with respect to the motion parameters. We do so by taking a first order Taylor expansion of the data residual  $r_d(\theta_s, o_x^k, x)$  around  $\hat{\theta}_s = \theta_s^k$ . Using eq. (5) this leads to

$$\tilde{r}_d(\theta_s, o_x^k, x) = \nabla_{\theta_s}^T r_d(\hat{\theta}_s, o_x^k, x) (\theta_s - \hat{\theta}_s) + r_d(\hat{\theta}_s, o_x^k, x). \quad (20)$$

This approximation is central to our estimation scheme. In particular the gradient of the data residual with respect to the motion parameters (i.e.  $\nabla_{\theta_s} r_d$ ) is the main ingredient

of the subsequent derivations. In Appendix A we derive it analytically, in terms of the motion model matrix  $m$  and the image gradient  $\nabla_x I$ . More specifically, in Appendix A, we show that

$$\nabla_{\theta} r_d(\theta, o_x^k, x) = m g_x(\theta, o_x^k, x). \quad (21)$$

where  $g_x(\theta, o_x^k, x)$  is a linear combination of the spatial image gradient in the previous and in the next frame, that is

$$g_x(\theta, o_x^k, x) = o_x^k \nabla_x I(x + m^T \theta, t + 1) + (1 - o_x^k) \nabla_x I(x - m^T \theta, t - 1) \quad (22)$$

Furthermore, in Appendix A, we show that eq. (20) is a form of the residual of the optical flow equation (i.e.  $(\nabla_x I)^T u + I_t$ ). More specifically, in Appendix A, we show that

$$\tilde{r}_d(\theta_s, o_x, x) = g_x(\hat{\theta}_s, o_x^k, x) u(\theta_s - \hat{\theta}_s, x) + r_d(\hat{\theta}_s, o_x^k, x) \quad (23)$$

### C. Optimization with respect to the motion parameters

Clearly, the optimization of eq. (3) with respect to the motion parameters of patch  $s$  (i.e.  $\theta_s$ ) keeping all the other parameters frozen is equivalent to the optimization of the local cost

$$L_s^{k+1}(\theta_s) = \lambda_d f_d(\theta_s, O_s^k, s) + \lambda_c \sum_{s' \in N_s} f_c(\theta_s, \theta_{s'}^k, s, s') \quad (24)$$

which, using eq. (4) and eq. (6),

$$= \lambda_d \sum_{x \in \Omega(s)} \rho(\tilde{r}_d(\theta_s, o_x^k, x)) + \lambda_c \sum_{s' \in N_s} b_{ss'} \rho(r_c(\theta_s, \theta_{s'}^k, \Omega(s, s'))) \quad (25)$$

where  $\tilde{r}_d(\theta_s, o_x, x)$  is as defined in eq. (20).

Subsequently, as in section III-A, we turn the optimization of eq. (25) into a weighted Least Squares problem by substituting, at iteration  $k + 1$ , the robust function  $\rho(r)$  by the quadratic function  $l(r^k)r^2$ . The equivalent weighted Least Squares problem that results is the optimization, with respect to  $\theta_s$ , of

$$\lambda_d \sum_{x \in \Omega(s)} l_d(\tilde{r}_d(\theta_s^k, o_x^k, x)) \left( \nabla r_d(\hat{\theta}_s, o_x^k, x) (\theta_s - \hat{\theta}_s) + r_d(\hat{\theta}_s, o_x^k, x) \right)^2 + \lambda_c \sum_{s' \in N_s} b_{ss'} l_c(r_c(\theta_s^k, \theta_{s'}^k)) (\theta_s - \theta_{s'}^k)^T M_{ss'} (\theta_s - \theta_{s'}^k) \quad (26)$$

The minimum of eq. (26) with respect to  $\theta_s$  is at the solution of the linear system

$$[A_d + A_c] \theta_s = B_d + B_c \quad (27)$$

with

$$A_d = \lambda_d \sum_{x \in \Omega(s)} l_d \nabla r_d \nabla^T r_d \quad (28)$$

$$A_c = \lambda_c \sum_{s' \in N_s} l_c b_{ss'} M_{ss'} \quad (29)$$

$$B_d = \lambda_d \sum_{x \in \Omega(s)} l_d \left[ \nabla r_d \nabla^T r_d \hat{\theta}_s - r_d \nabla r_d \right] \quad (30)$$

$$B_c = \lambda_c \sum_{s' \in N_s} l_c b_{ss'} M_{ss'} \theta_{s'}^k \quad (31)$$

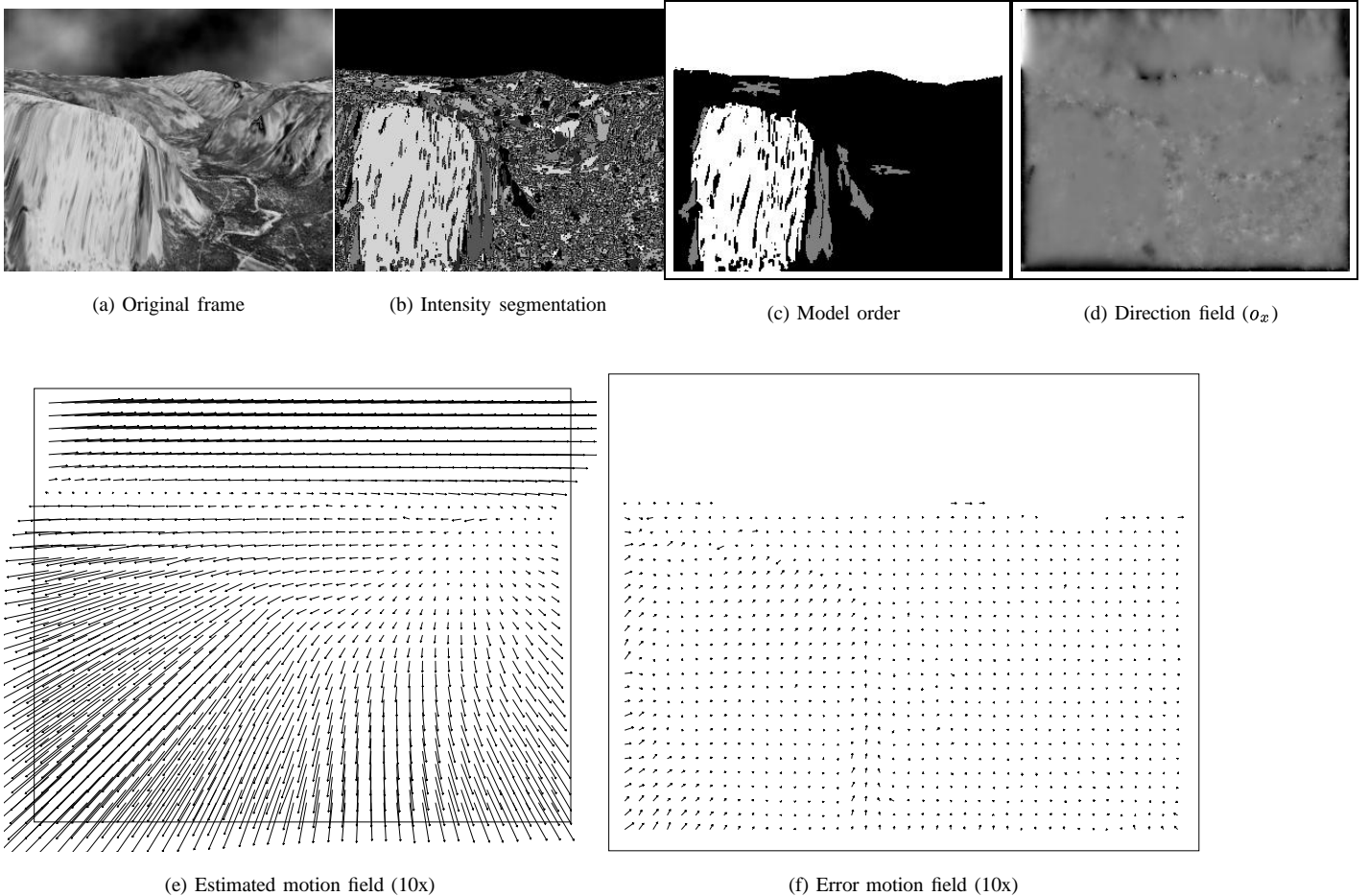


Fig. 2. “yosemite” sequence. In Fig. 2(c) a black value indicates a two parameter motion model, a grey indicates a four parameter motion model and white the full affine motion model. The initial intensity segmentation is obtained with no morphological simplification and  $T = 5$ .

where the functional dependencies of the terms that comprise eq. (28) - eq. (31) are omitted for notational simplicity. In the linear system of eq. (27) the role of the data and the motion regularization constraints is apparent. This linear system is straightforwardly solved, that is

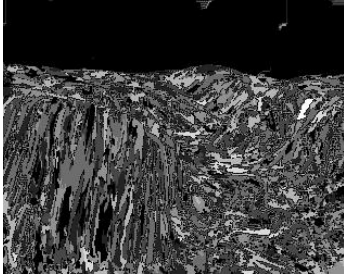
$$\theta_s^{k+1} = [A_d + A_c]^{-1} [B_d + B_c]. \quad (32)$$

where  $[A_d + A_c]^{-1}$  is the covariance matrix of our estimation. Models of variable order are treated by considering only the relevant rows and columns in eq. (32).

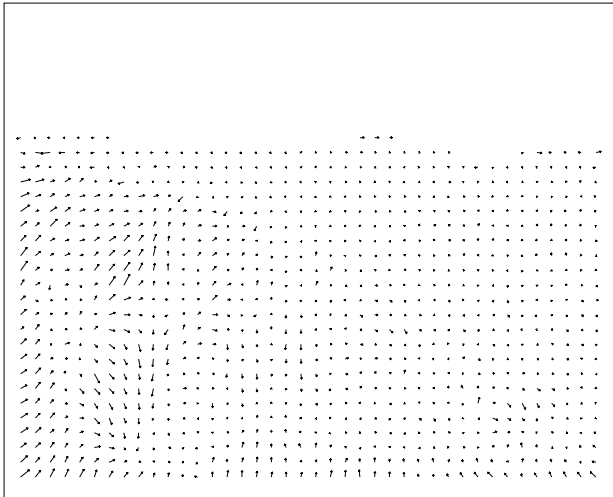
#### IV. EXPERIMENTAL RESULTS

We have applied our method in a number of synthetic and real image sequences and here we summarize the results. The synthetic image sequences are the well-known “yosemite” image sequence as well as a number of image sequences in which a rectangular is translated and rotated in a static background. We also present results for two real image sequences, namely the “jardin” image sequence, that contains complex articulated

human motion in a moving background and the “calendar” image sequence that contains rigid objects that translate and rotate in a moving background. In all of the experiments, patches were extracted with the segmentation scheme described at Section II-A. Typical segmentation results are depicted in Fig. 4 and Fig. 5, where the degree of oversegmentation is obvious. For all of the experiments (except of one experiment for the “yosemite” sequence) we used a  $5 \times 5$  square structuring element for morphological operations (opening and closing by reconstruction). The order of the models that were used for each patch were chosen depending on the dimensions of the patch in question. Segments with width (height) smaller than a threshold (set to 35 pixels throughout all of our experiments) are constrained to parametric models that are constant with respect to  $i(x)$  ( $i(y)$  respectively). For all of our experiments we kept the scale parameter  $\sigma_d$  of the data residual constant ( $\sigma_d = 10/\sqrt{2}$ ) and we adopt the linear reduction scheme of Black and Anandan [6] for the scale parameter  $\sigma_c$  of the motion residual between the values of  $1.5/\sqrt{2}$  and  $1/2\sqrt{2}$ . All the results were obtained with a 3 level multiscale framework with 20 iterations (an iteration here is a full cycle through the steps A, B and C of Table I) at each level except of the results obtained for the “jardin” image sequences for which 10 iterations were used at each level. Finally, for all the



(a) Intensity segmentation



(b) Error motion field (10x)

Fig. 3. “yosemite” sequence. Results obtained with an intensity segmentation using a  $5 \times 5$  structuring element for morphological simplification and  $T = 2$ .

experiments the motion parameters were initialized to zero (i.e.  $\theta_s = 0$ ) and the direction field at each pixel was initialized to 0.5 (i.e.  $o_x = 0.5$ ) so that at the beginning of the optimization procedure the backward and the forward constraints have equal significance. Typical execution times are around 90 seconds on a 2Ghz Pentium for a  $320 \times 250$  image.

The results for the “yosemite” image sequence are summarized in Fig. 2 and in Table II. In Fig. 2 we present the original frame, the initial intensity segmentation, the order of the parametric model of each patch, the estimated direction field, the estimated motion field and the error in the estimated motion field (omitting the sky region). The initial intensity-based segmentation was obtained without an initial morphological simplification and with  $T = 5$ . Note the degree of oversegmentation in Fig. 2(b) (a large number of patches contain single pixels) which clearly demonstrates the ability of the proposed method in dealing with patches of different sizes. The structure of the motion field is accurately recovered and there is little structure in the true error except around the image borders. It is also clear that the direction field is quite well estimated, with lower values at areas near the image borders that are visible only at the previous frame and average

values at the areas at the rest of the image that are visible at both frames. The true angular error (Table II, last row) remains low and comparable with the best results obtained so far by other researchers in the field<sup>1</sup>. We have obtained an average error (last row of Table II) which is lower than the method of Black and Jepson [5] that does not impose inter-patch constraints, but higher than the method of Memin and Perez that perform adaptive motion-based segmentation. Our method suffers mainly at the lower left corner where the initial intensity segmentation cannot produce larger patches. Therefore, at these areas, the benefits of the initial intensity segmentation are limited.

Technique		Average error	Std. deviation	Density
Szeliski & Coughlan [27]		2.45°	3.05°	100%
Szeliski & Shum [28]		2.20°	5.87°	23.1%
Black [4]		3.52°	3.25°	100%
Black & Anandan [6]		4.46°	4.21°	100%
Black & Jepson [5] (Parametric)		2.9°	3.2°	100%
Black & Jepson [5]		2.29°	2.25°	100%
Bab-Hadiashar & Suter [3]		1.97°	1.96°	100%
Memin & Perez [18]		1.58°	1.21°	100%
Patch Motion	sel = $5 \times 5$ , $T = 2$	2.73°	1.99°	100%
	$o_x = 1$ , $T = 5$	2.70°	2.13°	100%
	$o_x = 0$ , $T = 5$	2.37°	2.94°	100%
	$T = 5$	<b>2.24°</b>	<b>1.81°</b>	<b>100%</b>

TABLE II  
COMPARATIVE ANGULAR ERROR ON “YOSEMITE” SEQUENCE (WITHOUT SKY REGION)

For the same sequence, in Table II we summarize the results obtained for three additional number of settings. In order to illustrate the influence of the initial intensity segmentation, we present results obtained with the “standard” parameters for our initial intensity segmentation, that is with a  $5 \times 5$  structuring element for morphological simplifications and with  $T = 2$ . Such an initial intensity-based segmentation produces smaller patches for the facet of the mountain (Fig. 3(a)) which results in a larger angular error at that area (Fig. 3(b)) in correspondence to the one obtained when a large patch covers most of facet of the mountain (Fig. 2(b)). This seems to be the main reason for a relatively higher average angular error (row 9, Table II). In order to illustrate the influence of the direction field we summarize the results obtained with fixed values of the direction field that correspond either to backward only ( $o_x = 0$ ) or to forward only ( $o_x = 1$ ) motion estimation. As expected, since the “yosemite” is a sequence with zoom and thus with occlusions mainly in the next frame, the worst performance is obtained with the forward only motion estimation while the backward only motion estimation produced comparable results (rows 10-11, Table II).

We have also generated six synthetic image sequences namely r1-r4 and t1-t2 in which we translate/rotate a rectangle in a static background. The rectangles contain different degree of texture. In sequences r1-r4 we use as the rectangle a frame from the “rubic” image sequence, which has large areas with low intensity variation and in sequences t1 and t2 we use a

<sup>1</sup>More results can be found in [22]

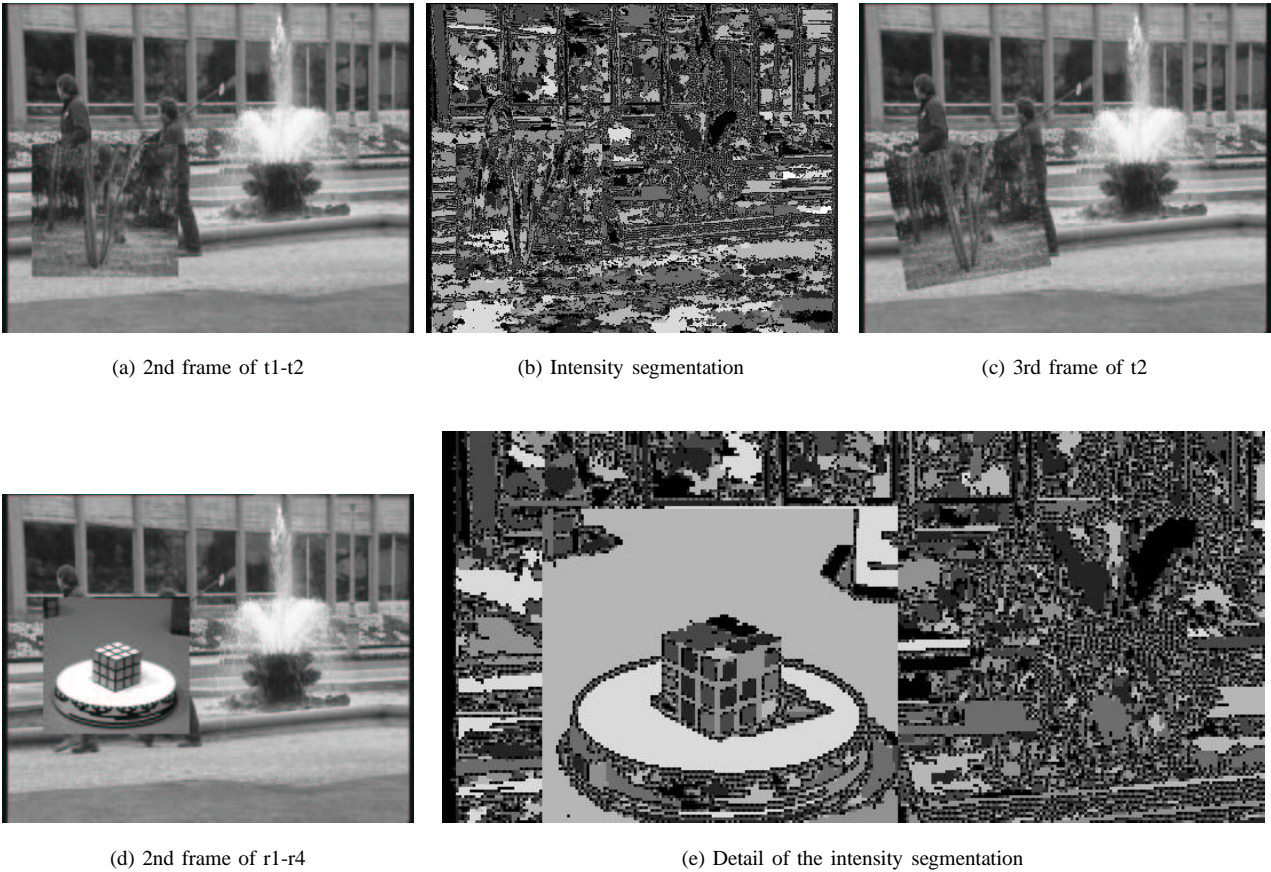


Fig. 4. “Synthetic” image sequences: Translating/rotating rectangulars

frame for the “trees” image sequence. The first frames for each group of sequences, the corresponding initial intensity segmentations and the second frame of sequences r4 and t2 are depicted in figure 4. In Table III we give the corresponding translation and rotation parameters.

Sequence	r1	r2	t1	r3	r4 & t2
$(u, v, \phi)$	$(5, 2, 0^\circ)$	$(10, 2, 0^\circ)$	$(5, 1, 0^\circ)$	$(5, 2, 5^\circ)$	$(5, 2, 10^\circ)$

TABLE III

TRANSLATION AND ROTATION PARAMETERS FOR THE SYNTHETIC IMAGE SEQUENCES

The results for the synthetic image sequences are summarized in Table IV where the average error, the standard deviation and the percentage of motion vectors whose angular error is smaller than certain thresholds are presented for the proposed method and the pixel-based method presented in [6]. The proposed method consistently outperforms the method in [6]. More interestingly, the larger differences in the performance are recorded for the sequences r1-r4 which contain large areas with low intensity variation. In the proposed scheme, such areas form large patches (fig. 4(e)) which makes it possible to derive data constraints that can support the estimation of the few parameters of their motion model. In comparison, the data constraints of pixel-based methods are completely unreliable within such areas.

Seq.	mean	$\sigma$	Vectors with error less than (%)				
			1°	2°	3°	5°	10°
r1	0.82°	7.05°	95.76	97.43	98.17	98.58	98.86
r2	1.45°	9.96°	94.70	96.22	96.93	97.44	98.15
r3	0.71°	4.82°	86.80	94.13	96.77	98.50	99.34
r4	1.87°	7.58°	83.27	86.88	88.78	90.88	94.86
t1	0.30°	4.07°	98.43	98.71	98.94	99.29	99.44
t2	0.91°	6.09°	93.71	94.56	95.19	95.84	97.48
r1 [6]	7.29°	16.99°	77.14	80.63	82.35	83.76	84.72
r2 [6]	5.00°	12.87°	75.63	79.27	81.23	82.94	84.24
r3 [6]	2.03°	7.93°	78.05	82.74	86.40	91.96	96.45
r4 [6]	5.46°	15.78°	77.60	80.78	82.70	84.90	88.00
t1 [6]	2.06°	8.64°	91.54	92.36	92.77	93.25	93.51
t2 [6]	1.52°	7.99°	91.26	92.44	93.15	94.25	96.59

TABLE IV

COMPARATIVE RESULTS FOR SYNTHETIC IMAGE SEQUENCES: ANGULAR ERROR

Note that the synthetic image sequences contain both areas that are occluded at the next frame and areas that are occluded at the previous frame. In order to illustrate the positive influence of the direction field in such a setting, we provide in Table V comparative results with forward only motion estimation (i.e.  $o_x = 1$ ). It is clear that the backward/forward motion estimation consistently outperforms the one direction motion estimation especially for the sequences that the motion is rather large in magnitude.

In order to demonstrate the behavior of our method in an image sequence with a more challenging motion we



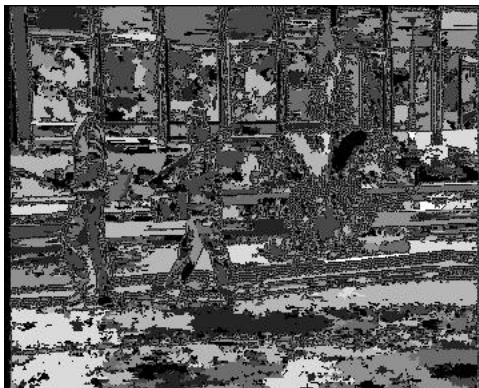
Sequence	r1	r2	t1
$(\mu, \sigma)$ full method	$(0.82^\circ, 7.05^\circ)$	$(1.45^\circ, 9.96^\circ)$	$(0.30^\circ, 4.07^\circ)$
$(\mu, \sigma), o_x = 1$	$(0.81^\circ, 7.47^\circ)$	$(1.76^\circ, 11.63^\circ)$	$(0.48^\circ, 5.84^\circ)$
Sequence	r3	r4	t2
$(\mu, \sigma)$ full method	$(0.71^\circ, 4.82^\circ)$	$(1.87^\circ, 7.58^\circ)$	$(0.91^\circ, 6.09^\circ)$
$(\mu, \sigma), o_x = 1$	$(0.81^\circ, 5.00^\circ)$	$(2.65^\circ, 13.16^\circ)$	$(1.01^\circ, 7.03^\circ)$

TABLE V

COMPARATIVE RESULTS WITH THE FULL METHOD (I.E. WITH ESTIMATION OF  $o_x$ ) AND WITH FORWARD ONLY MOTION ESTIMATION (I.E. WITH A FIXED DIRECTION FIELD  $o_x = 1$ ) FOR THE SYNTHETIC SEQUENCES



(a) Original frame



(b) Intensity Segmentation

Fig. 5. “Jardin” sequence. Two walking men in a moving background

present results for the image sequence “Jardin” that depicts two walking men in a moving background. In Fig. 5 we present the first frame of the sequence and the corresponding intensity segmentation while Fig. 6 depicts the motion field and its horizontal and vertical components as these were estimated by the proposed method. It is clear that the motion field is estimated rather well, the discontinuities are well preserved and, although the size of the patches vary considerably (Fig. 5(b)), the motion field is not noisy. The largest inaccuracies take place at the area between the first man and the left border of the image in which there is not much information about the horizontal motion, and at very

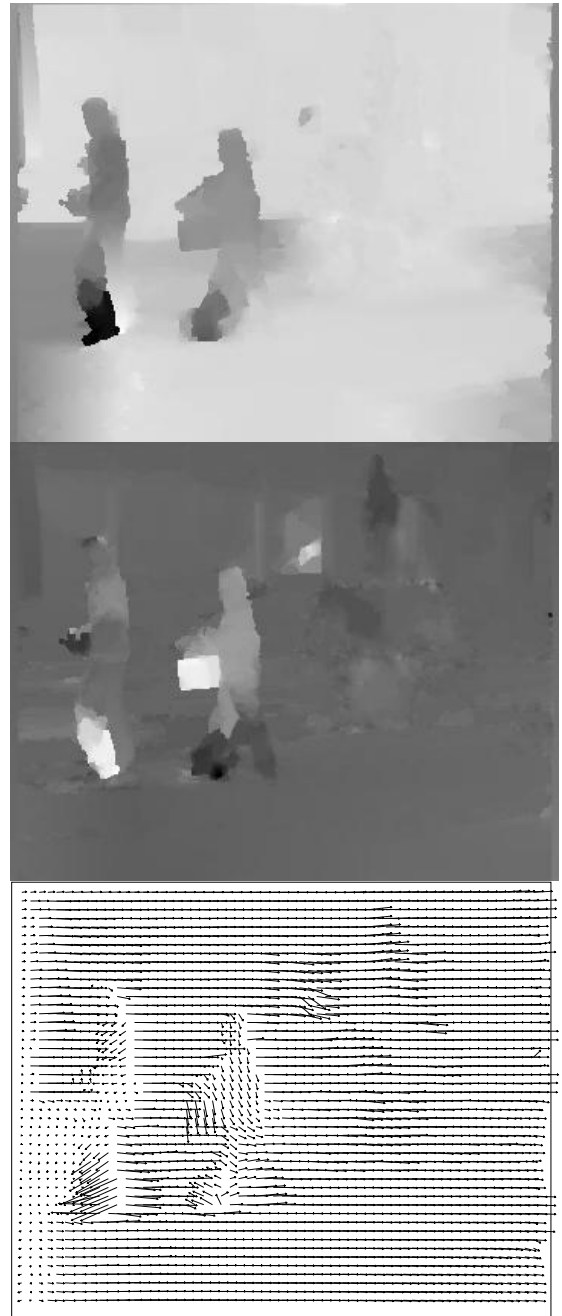


Fig. 6. Results for “Jardin” sequence with the proposed method. First row: Horizontal motion component. Second row: Vertical motion component. Third row: Full flow.

thin areas such as the stick of the broom that the second man carries. Furthermore, the “textured” motion of the water of the fountain is oversmoothed. However, our method constitutes an improvement over pixel-based methods such as [6] (Fig. 7). Indeed, it is clear that the discontinuities are better preserved, the motion of the legs of the two men is better recovered and the motion field is in general less noisy. Although the noise in the motion field estimated by the pixel-based method [6] can be probably reduced by applying the patch-based improvement described in [5], the motion of the legs of the first man would never be recovered since the initialization is far from the

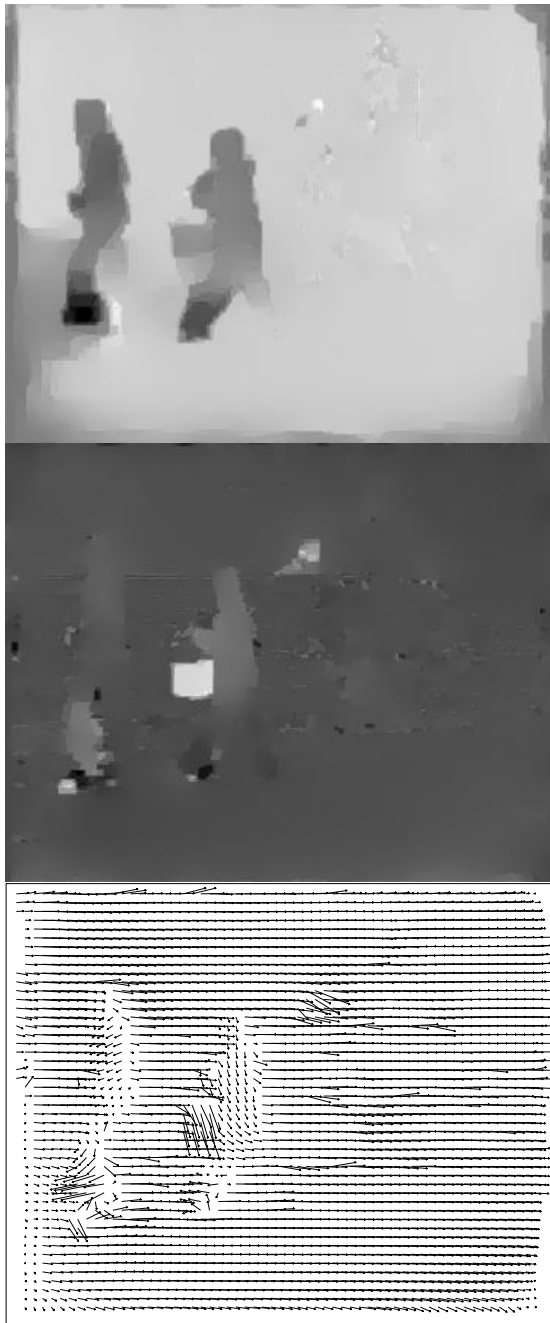


Fig. 7. Results for “Jardin” sequence with the method of Black [4]. First row: Horizontal motion component. Second row: Vertical motion component. Third row: Full flow.

correct motion field. These are better illustrated in Fig. 8, in which we present details of the horizontal motion component (first row, Fig. 6 and Fig. 7). Finally, in Fig. 8(c) we present a detail of the estimated direction field and in Fig. 8(d) a detail of the the horizontal motion component that is estimated for a fixed direction field ( $o_x = 0$ ). Although the direction field is not perfectly estimated, in rough lines it corresponds

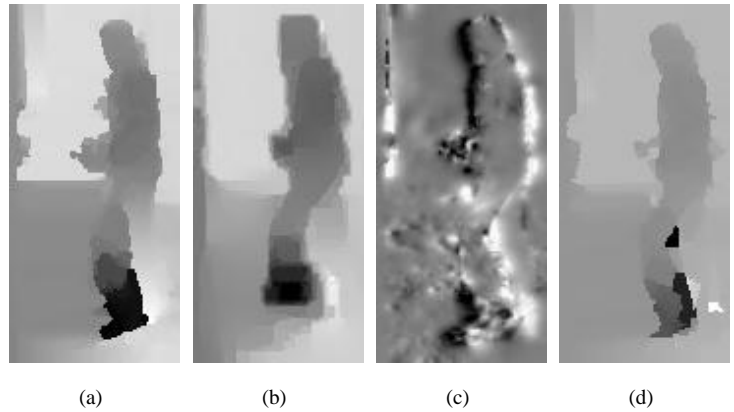


Fig. 8. “Jardin” sequence. (a)-(b) detail of the horizontal component of the estimated motion field (c) detail of the estimated direction field  $o_x$  and (d) detail of the horizontal component of the estimated motion field for backward only estimation (with fixed  $o_x = 0$ ).

to the direction at which the true correspondences lie<sup>2</sup>. In comparison, when the direction field is fixed ( $o_x = 0$ ) the motion estimation is clearly deteriorated at areas that are occluded at the previous frame, such as the area behind the leg of the man.

Finally, in Fig. 9 we summarize the results for the “calendar” image sequence in which a number of rigid objects translate/rotate in front of moving background. Our method was able to recover well the structure of the motion of the scene and at the same time localize quite well the motion discontinuities. The most obvious errors are in the area in front of the ball which could be probably due to its shadow and at the thin area under the train where the motion field seems oversmoothed.

## V. CONCLUSIONS

In this paper we have presented a method for the estimation of dense motion fields which is based on the application of strong parametric constraints on the motion field within image patches and weak smoothness constraints on the motion field along the edges of neighboring patches. We have expressed the motion estimation as an optimization problem and solved it in an iterative scheme for the motion parameters of the patches.

The advantages of the proposed method can be summarized as follows:

- The initial intensity segmentation groups in advance areas with low intensity variation into larger patches that exhibit larger intensity variation and which, in general, yield more reliable data constraints. Furthermore, by utilizing the initial intensity segmentation we enforce that in the estimated motion field motion discontinuities (if any) will coincide with intensity edges.
- Motion estimation and regularization at patch level are expressed in a single framework. Our framework is

<sup>2</sup>It is also interesting to note that it is rather textured at areas that the motion estimation is not very accurate, especially if a small regularization coefficient  $\lambda_o$  is chosen.

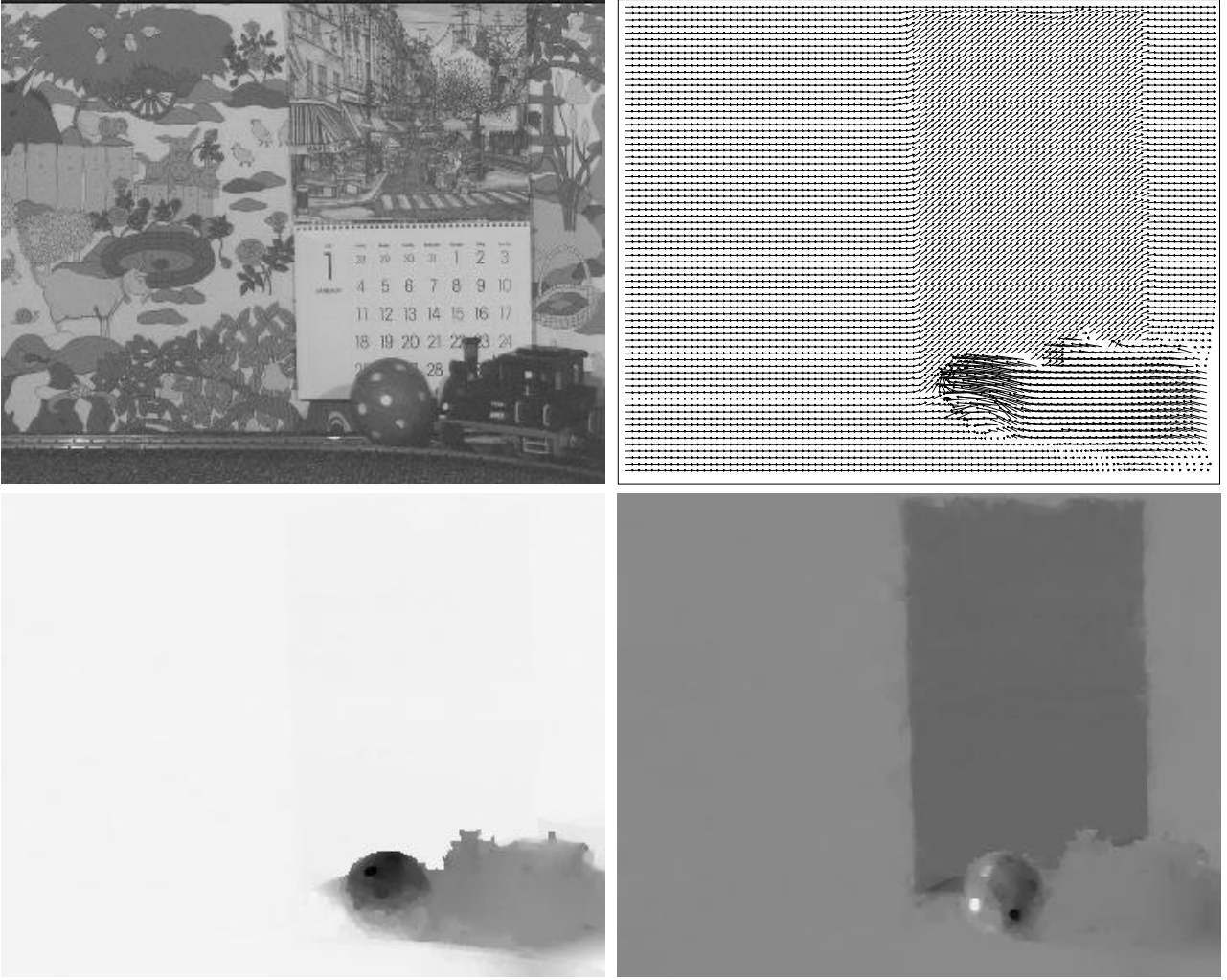


Fig. 9. “calendar” sequence: From left to right and top to bottom: Original frame, estimated motion field, horizontal motion component and vertical motion component.

general enough to contain pixel-based motion estimation schemes and region-based motion estimation schemes as special cases. Moreover, our formulation uses parametric models of different order for each patch, depending on their size and shape, and can deal successfully with patches of different, even very small, size.

Results were presented for both synthetic and real image sequences. We were able to obtain accurate piecewise smooth motion fields in the presence of motion large in magnitude and motion discontinuities.

## VI. APPENDIX A

In this appendix we will derive eq. (21) and eq. (23).

Eq. (5) defines the data residual at  $x$  as a linear combination of the forward and backward data residuals. Let us derive the gradient of the forward data residual  $r_d^f$ . In order to do so, we will change slightly the notation, so that it is a function of  $u(\theta)$ , that is  $r_d^f(u(\theta, x), x)$ . Then,

$$r_d^f(u(\theta, x), x) = I(x + u(\theta, x), t + 1) - I(x, t) \quad (33)$$

implies that

$$\nabla_{\theta} r_d^f(u(\theta, x), x) = \nabla_{\theta} (I(x + u(\theta, x), t + 1) - I(x, t)) \quad (34)$$

$$= \nabla_{\theta} I(x + m^T \theta, t + 1) \quad (35)$$

$$= m \nabla_x I(x + m^T \theta, t + 1). \quad (36)$$

Similarly the gradient of the backward data residual is given as,

$$\nabla_{\theta} r_d^b(u(\theta, x), x) = m \nabla_x I(x - m^T \theta, t - 1) \quad (37)$$

Then, the gradient of the data residual (eq. (5)), is a linear combination of the gradients of  $r_d^f$  (eq. (36)) and  $r_d^b$  (eq. (37)), that is

$$\nabla_{\theta} r_d(u(\theta, x), o_x^k, x) \quad (38)$$

$$= o_x^k \nabla_{\theta} r_d^f(u(\theta, x), x) + (1 - o_x^k) \nabla_{\theta} r_d^b(u(\theta, x), x) \quad (39)$$

$$= o_x^k m \nabla_x I(x + m^T \theta, t + 1) + (1 - o_x^k) m \nabla_x I(x - m^T \theta, t - 1) \quad (40)$$

Rearranging the terms of eq. (40) and defining  $g_x$  according to eq. (22) we arrive at eq. (21), which concludes the proof.

Under this expansion it is clear how to construct the linear system of eq. (27) which is the core of our estimation scheme. Note that eq. (21) is derived for a general motion model matrix  $m$ . In order to make the derivation more clear we derive analytically the gradient of the data residual (with respect to the motion parameters) in the case of the affine motion model. We do so by substituting the affine motion model matrix  $m$  (eq. (2)) in eq. (21). That is,

$$\begin{aligned} \nabla_{\theta}^T r_d(u(\theta, x), x) \\ = [g_{i(x)} \ i(x)g_{i(x)} \ j(x)g_{i(x)} \ g_{j(x)} \ i(x)g_{j(x)} \ i(y)g_{j(x)}] \end{aligned} \quad (41)$$

where  $(g_{i(x)}, g_{j(x)})^T \equiv g_x$ .

In order to derive eq.23 we simply substitute eq.21 in eq.20. Then,

$$\tilde{r}(\theta_s, x) = \nabla_{\theta_s}^T r_d(u(\hat{\theta}_s, x), x)(\theta_s - \hat{\theta}_s) + r_d(u(\hat{\theta}_s, x), x) \quad (42)$$

$$= (mg_x)^T(\theta_s - \hat{\theta}_s) + r_d(u(\hat{\theta}_s, x), x) \quad (43)$$

$$= g_x^T(m^T(\theta_s - \hat{\theta}_s)) + r_d(u(\hat{\theta}_s, x), x) \quad (44)$$

$$= g_x^T u(\theta_s - \hat{\theta}_s, x) + r_d(u(\hat{\theta}_s, x), x) \quad (45)$$

## REFERENCES

- [1] L. Alvarez, J. Weickert, and J. Sanchez. Reliable estimation of dense optical flow fields with large displacements. *Int'l Journal of Computer Vision*, 39(1):41–56, Aug. 2000.
- [2] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *Int'l Journal of Computer Vision*, 2:283–310, 1989.
- [3] A. Bab-Hadiashar and D. Suter. Robust optic flow computation. *Int'l Journal of Computer Vision*, 29(1):59–77, Aug. 1998.
- [4] M. Black. Recursive on-linear estimation of discontinuous flow fields. In *Proc. Europ. Conf. Computer Vision*, pages 138–145, 1994. Stockholm, Sweden.
- [5] M. Black and A. Jepson. Estimation optical flow in segmented images using variable-order parametric models with local deformation. *IEEE Trans. PAMI*, 18(10):972–986, Oct. 1996.
- [6] M.J. Black and P. Anandan. The robust estimation of multiple motions: Parametric and piecewise-smooth flow-fields. *Computer Vision and Image Understanding*, 63(1):75–104, Jan. 1996.
- [7] M.J. Black and A. Rangarajan. On the unification of line processes, outlier detection and robust statistics with applications in early vision. *Int'l Journal of Computer Vision*, 19(1):57–91, Jan. 1996.
- [8] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, Cambridge MA, 1987.
- [9] V. Dang, A. Mansouri, and J. Konrad. Motion estimation for region-based video coding. In *Int. Conf. Image Processing*, pages 189–192, 1995. Washington DC, USA.
- [10] L. Gaucher and G. Medioni. Accurate motion flow estimation with discontinuities. In *Int'l Conf. Computer Vision*, pages 695–702, Sept. 1999. Corfu, Greece.
- [11] M. Gelgon and P. Bouthemy. A region-level motion-based graph representation and labeling for tracking a spatial image partition. *Pattern Recognition*, 33:725–740, Apr. 2000.
- [12] S. Geman and D. Geman. Stochastic relaxation, gibbs distribution and the bayesian restoration of images. *IEEE Trans. PAMI*, 6:721–741, Nov. 1984.
- [13] F. Heitz and P. Bouthemy. Multimodal estimation of discontinuous optical flow using markov random fields. *IEEE Trans. PAMI*, 15(12):1217–1232, 1993.
- [14] P.W. Holland and R.F. Welsch. Robust regression using iteratively reweighted least squares. *Comm. Statistics. Theory and Methods*, 1977.
- [15] B. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17(1-3):185–203, Aug. 1981.
- [16] P. Huber. *Robust Statistics*. John Willey & Sons, 1981.
- [17] M. Irani, R. Rousso, and S. Peleg. Detecting and tracking multiple moving objects using temporal integration. In *ECCV*, pages 282–287, 1992.
- [18] E. Memin and P. Perez. Hierarchical estimation and segmentation of dense motion fields. *IJCV*, 46(2):129–155, February 2002.
- [19] David W. Murray and Bernard F. Buxton. Scene segmentation from visual motion using global optimization. *IEEE Trans. Pattern Analysis and Machine Intelligence*, PAMI-9(2):220–228, Mar 1987.
- [20] H.H. Nagel and W. Enkelmann. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8(5):565–593, Sep. 1986.
- [21] J. M. Odobez and P. Bouthemy. Robust multiresolution estimation of parametric motion models. *Journal of Visual Communication and Image Representation*, 6(4):348–365, Dec. 1995.
- [22] E.P. Ong and M. Spann. Robust optical flow computation based on least-median-of-squares regression. *Int'l Journal of Computer Vision*, 31:51–82, Feb. 1999.
- [23] I. Patras, E.A. Hendriks, and R.L. Legendijk. Video segmentation by map labeling of watershed segments. *IEEE Trans. PAMI*, 22(3):326–332, Mar. 2001.
- [24] I. Patras and M. Worring. Patch motion estimation. In *Int'l Conf. Pattern Recognition*, volume 1, pages 323–326, Aug. 2002. Quebec City, Canada.
- [25] C. Stiller. Object-based estimation of dense motion fields. *IEEE Trans. Image Processing*, 6(2):234–250, Feb. 1997.
- [26] C. Stiller and J. Konrad. Estimating motion in image sequences: A tutorial on modeling and computation of 2d motion. *IEEE Signal Processing Magazine*, 16:70–91, Jul. 1999.
- [27] R. Szeliski and J. Coughlan. Spline-base image registration. Technical Report CRL 94/1, Digital Equipment Corporation, 1994.
- [28] R. Szeliski and H.-Y. Shum. Motion estimation with quadtree splines. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18(12):1199–1210, Dec. 1996.
- [29] Y. Weiss and E.H. Adelson. A unified mixture framework for motion segmentation: Incorporating spatial coherence and estimating the number of models. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, pages 321–326, 1996.